Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 16 (2006) 1014–1020

3D-QSAR studies on antitubercular thymidine monophosphate kinase inhibitors based on different alignment methods

V. Aparna,[a] J. Jeevan,[a] M. Ravi,[a] G. R. Desiraju[a,*] and B. Gopalakrishnan[b,*]

[a]*School of Chemistry, University of Hyderabad, Hyderabad 500 046, India*
[b]*Life Sciences R&D Division, Advanced Technology Centre, TATA Consultancy Services Limited, 1 Software Units Layout, Madhapur, Hyderabad 500 081, India*

Received 8 July 2005; revised 7 October 2005; accepted 25 October 2005
Available online 15 November 2005

Abstract—Three dimensional quantitative structure–activity relationship (3D-QSAR) studies were carried out on deoxythymidine monophosphate (dTMP) derivatives inhibiting thymidine monophosphate kinase (TMPK) in *Mycobacterium tuberculosis*. Molecular field analysis (MFA) models with three different alignment techniques, namely, least squares, pharmacophore based and receptor based methods were developed. Receptor based MFA model showed better results when compared with least squares and pharmacophore based models. The results help us to understand the nature of substituents required for activity and thereby provide guidelines to design novel and potent inhibitors as antitubercular agents.

Tuberculosis (TB) is alarmingly on the rise.[1] Approximately, one third of the world's population is infected with TB bacillus, *Mycobacterium tuberculosis*, with more than 8 million people contracting the disease and 2 million people dying of it each year.[2] A peculiar aspect of its pathogenicity comes from the fact that it can remain quiescent and become active decades later. One of the most significant risk factors for developing tuberculosis is human immunodeficiency virus (HIV) infection.[3] The World Health Organisation (WHO) reports that in Africa 80% of the cases with TB are HIV infected. The current treatment of active TB includes a dosage regime of four drugs (isoniazid, rifampicin, pyrazinamide, and ethambutol) for at least six months. As a consequence of the prolonged duration, irregular treatment, and the highly adaptive nature of the organisms to their surroundings, multidrug resistant (MDR) strains of *M. tuberculosis* have developed. Therefore, the urgent need to cope with the current crisis has stimulated the search for new targets and the development of new antibiotic drugs to meet the global emergency.

M. tuberculosis thymidine monophosphate kinase ($TMPK_{mt}$) belongs to a large super family of nucleoside monophosphate kinases (NMPK). It catalyses the phosphorylation of deoxythymidine monophosphate (dTMP) to deoxythymidine diphosphate (dTDP) utilizing ATP as a phosphoryl donor. This step lies at the junction of the de novo and salvage pathways of thymidine triphosphate (TTP) metabolism and is the last specific enzyme for its synthesis.[4] Also, the sequence of $TMPK_{mt}$ when compared with that of its human isozyme shows only 22% sequence identity.[5] These characteristics make $TMPK_{mt}$ one of the potential targets for the design of new antitubercular drugs.

Several dTMP derivatives were synthesized and studied for their effect on the $TMPK_{mt}$.[6–9] Structure–activity relationship studies on these reveal that a halogen substitution at position-5 reflects the size effect, with the halogen serving as a cavity filler.[10] In the present study, 3D-QSAR analyses were carried out on a series of dTMP analogues, to gain further insight into the key structural features required to design potential drug candidates of this class. The fact that the reliability and

Keywords: Thymidine monophosphate kinase; 3D-QSAR; Alignment; Antitubercular agents.
* Corresponding authors. Tel.: +91 40 55673591; fax: +91 40 55672222 (B.G.); tel.: +91 40 23134828; fax: +91 40 23010567 (G.R.D.); e-mail addresses: gautam_desiraju@yahoo.com; gopal@atc.tcs.co.in

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.10.086

the efficiency of 3D-QSAR methods depend on the orientation of the molecule[11] prompted us to generate and compare molecular field analysis (MFA) models using different alignment methods.

Selection of molecules. A data set of 47 compounds reported as thymidine monophosphate kinase inhibitors was collected from the literature (Table 1).[6–9] The inhibitory activities were converted into the corresponding pK_i ($-\log K_i$) values, where the K_i value represents the drug concentration that causes inhibition of dTMP phosphorylation by $TMPK_{mt}$. All the K_i values had been obtained by the same assay method.[7] pK_i values of all the molecules spanned a sufficiently wide range from 2 to 6. The training/test set selection was done manually such that they populate the wide range of activities in similar proportions.

All the molecules were built from the coordinates of dTMP structure (1G3U) using the builder module of Cerius^2.[12] All the structures were minimized using the steepest descent algorithm with a convergence gradient value of 0.001 kcal/mol. Partial atomic charges were calculated using the Gasteiger method. Further geometry optimization was carried out for each compound with the MOPAC 6 package using the semi-empirical AM1 Hamiltonian.

Three different alignment methods that were used for MFA are detailed as follows.

Model I (least squares alignment): This alignment was carried out using the shape reference method in the QSAR module within Cerius^2 using the six thymidine ring atoms. The most active compound **38** was used as a template for superposing the rest of the molecules. The aligned molecules are shown in Figure 1a. From the 47 compounds above, 35 were used in the training set and the remaining 12 compounds were included in the test set.

Model II (pharmacophore based alignment): A pharmacophore model was developed using the CATALYST[13] HypoGen module (published elsewhere).[14] The studied molecules were superimposed on the best pharmacophore model using the 'best fit' method. The conformation of each molecule is selected based on the lowest rmsd between the pharmacophoric features and the corresponding functional groups present in the molecule. Figure 1b shows the alignment obtained. Out of the 47 molecules, 35 molecules were used as training set and the remaining 12 molecules were selected as test set.

Model III (receptor based alignment): Docking was carried out using the program GOLD.[15] The details of the docking setup are illustrated in Ref. 14. The docking run produces ten conformations for each molecule. Out of these, the best docking conformation for each molecule is selected based on GOLD score. Figure 1c shows the overlay diagram of the ligand conformations within the binding pocket of the $TMPK_{mt}$. Thirtyfour of fourtyseven molecules were selected as the training set leaving the other 13 molecules as test set.

MFA. Alternate probes like NH_2 and H_2O were used to calculate the electrostatic interaction energy. The geometry of H_2O probe was optimized at the 6-31G* level using Gaussian 03 and the corresponding charges were used for the calculation of interaction energies. In the case of the NH_2 probe, an arginine molecule was taken as template and optimized at the 6-31G* level. The charges corresponding to the side chain guanidinium NH_2 of arginine were considered, while adjusting the total charge of the probe to +0.1.

MFA studies were performed with the QSAR module of Cerius^2. The molecular field was created using CH_3, NH_2 and H_2O as probes representing steric and electrostatic fields, respectively. The steric and electrostatic fields were sampled at each point of a regularly spaced grid of 1 Å. A number of spatial and structural descriptors such as polarizability, dipole moment, radius of gyration, number of rotatable bonds, molecular volume, principle moments of inertia, A$\log P$, number of hydrogen bond donors and acceptors, and molar refractivity were considered along with the steric and electrostatic descriptors. Only 10% of the total descriptors whose variance was highest were considered for further analysis. Regression analysis was carried out using genetic partial least squares (G/PLS) method consisting of over 50,000 generations with a population size of 100. The optimal number of components was set to eight based on better r^2 and r^2_{cv} values for a given training set. An energy cutoff of ±30.0 kcal/mol was set for both steric and electrostatic contributions. The smoothing parameter, d, was set to 1.0 to control the bias in the scoring factors between equations with different number of terms. The length of the final equation was fixed to seven descriptors. The linear option was used in the equation creation. Cross-validation was performed with the leave-one-out procedure. The PLS analysis was scaled, with all variables normalized to a variance of 1.0.

Use of alternate probes for calculating electrostatic descriptors. MFA was carried out initially with the generic probe, (H^+) representing the electrostatic interactions. However, the interaction energies between (H^+) and all the phosphate containing molecules were too high above the cutoff value, overemphasizing the electrostatic energy component. The activities of these molecules were therefore predicted incorrectly. To eliminate this artifact, NH_2 and H_2O were used as alternate probes, as outlined in the methods section above, to appropriately sample the electrostatic environment. The choice of these probes was appropriate because of the presence of seven positively charged residues (Arg14, Arg74, Arg95, Arg153, Arg156, Arg160, and Lys13) as well as a number of bound water molecules (Wat1002, Wat1003, Wat1009, Wat1012, Wat1014, Wat1018, Wat1022, Wat1024, Wat1026, and Wat1050) in the active site of $TMPK_{mt}$.